



Developing A Streaming Media Capture And Delivery Service: The Experience at Hewlett Packard Labs

Jamie Dinkelacker, Mark Yoshikawa, Alan M. Parry
Research Operations
HP Laboratories Palo Alto
HPL-2003-96 (R.1)
May 5th, 2003*

E-mail: {jamie, marky, alan.parry} @hpl.hp.com

streaming
media, video,
digital
capture

Like many large, distributed organizations, Hewlett Packard Labs has facilities around the globe. As part of a research environment, it's common for specialists to give talks on current work, and to invite guest speakers on topics of interest. However, special expertise is also distributed within the geographic areas and time zones in which people live and work, so how can their talks be made available across time and distance? To this end, we've created digital encoding kiosks that can accept either video tapes (PAL, VHS) or live feeds from a camera, digitize these, and store them on a media server for subsequent viewing via streaming media on Windows or Linux platforms. Described below are the techniques used at HP Labs to provide this service.

Developing A Streaming Media Capture And Delivery Service: The Experience at Hewlett Packard Labs

Jamie Dinkelacker, Ph.D., Mark Yoshikawa, Alan M. Parry

Hewlett Packard Laboratories

Palo Alto, California USA and Bristol, England UK

Email: {jamie, marky, alan.parry} @ hpl.hp.com

April 2003

Abstract

Like many large, distributed organizations, Hewlett Packard Labs has facilities around the globe. As part of a research environment, it's common for specialists to give talks on current work, and to invite guest speakers on topics of interest. However, special expertise is also distributed within the geographic areas and time zones in which people live and work, so how can their talks be made available across time and distance? To this end, we've created digital encoding kiosks that can accept either video tapes (PAL, VHS) or live feeds from a camera, digitize these, and store them on a media server for subsequent viewing via streaming media on Windows or Linux platforms. Described below are the techniques used at HP Labs to provide this service.

Introduction

The modern, distributed organization faces numerous challenges in regard to getting information to the right people at the right time. HP Labs, like many research organizations or endeavors, is distributed around the world. It's commonplace to have people report on their research as well as host guest speakers from industry and academe. One significant problem is making these talks available to people distributed in time and space. Another is archiving and indexing them in such a way that they can be found and reviewed at some subsequent time, should the topic become important to someone who was either not at the organization at the time, or became interested in the topic at a later date.

At the time of this writing (April 2003) Hewlett Packard Laboratories is similar to many large and complex companies that have many events occurring. Not only do people work in various locations, they do not necessarily have time to attend them all -- but even learning about the talks and their schedule is a challenge. Also, our research covers numerous disciplines - computer science, electrical engineering, color science, mathematics, economics,

and so on. It's common for specialists to give talks on current research and to also invite guest speakers to address topics of interest. Making these talks available across time and distance to colleagues around the world is a challenge. To this end we developed "capture kiosks" for both capturing live presentations (video) or converting video tapes of talks (either VHS/NTSC or PAL) in to a streaming digital format.

The kiosks are used for generating streaming video in two ways: First, the kiosks are self-serve and enable anyone to convert a videotape into a streaming media format which is automatically stored on and registered with the appropriate servers for streaming. Although notifications of new programs are not automatic, it's not unusual for our Communications department to send out such announcements following significant events that have been digitized and are now available. Second, a kiosk setup can be used for live capture of an event by taking a feed from a camera and/or mixing board, converting it into a stream-able format, store it and then registering it with a streaming server. However, the thrust of this paper is predominantly focused on the self-serve encoding kiosks that are used for encoding recorded video tapes such as those that are typically done for archival purposes when talks and presentations are given.

This paper describes the configuration of the kiosks and servers, the desktop and laptop client programs deployed, our storage challenges, and how the service is being currently used.

The Capture Kiosks

The capture kiosks are self-serve, stand alone PC systems that can be used in either of two ways: (1) an operator can employ the kiosk to digitally encode a video tape into a streaming media format; or (2) a kiosk can also be on wheels and taken into a room where a live talk is being given. In this circumstance a feed can be taken from the camera and fed directly into the digitizing component of the kiosk. Previous efforts in this area within HP Labs have shown that the major challenge with such systems is to provide an interface that is both user friendly and accessible to non-technical employees. Moreover, this must cover both sides of the process - simplicity in encoding the media as well as ease of finding, retrieving, and playing back the video streams.

One of the benchmarks of system success is the depth and breadth of the media available in the library. To this end there was a need that the capture kiosks be operated with the least technical knowledge possible, otherwise they would essentially go underutilized. Digital media, and in particular digital video, is a morass of standards, settings, compression options, codec choices and more. Many non-technical users find this a barrier to entry when using video encoding systems. Wherever possible the technical choices are minimized for the operators and when that cannot be done, they are presented in terms of business decisions. An example of this is that the system is able to encode at higher bit rates if the media is required to be incorporated in live presentations or CD-ROMs as opposed to streamed over an intranet. The operator, with this decision factor, can then inform the kiosk system prior to encoding whether the content will be used in presentations or CD-

ROMs. If this is the case, the encoding bit rate is set programmatically on their behalf to the appropriate level. For example, for intranet playback that will probably be viewed on a PC screen in someone's cubicle, the typical encoding are audio streams at 12 Kbps and video at either 16 Kbps or at 148 Kbps. In contrast, for CD-ROM playback, a higher encoding density is used which is 32 Kbps for audio and 318.2 Kbps (30 frames per second). The technical details of the kiosks are below.

Hardware. Each capture kiosk is based on an HP xw8000 PC¹ with an Osprey 500 DV Pro² video board running Windows XP. Streaming video data is made available primarily to users of personal productivity computers (notebooks and desktops) that run either XINE³ or mplayer⁴ on Linux or Windows Media Player⁵. The benefit of the Osprey 500 DV Pro video board is that it provides custom hardware support for the encoding of Windows Media content, enabling the streaming video encoding process to take place real-time, removing the need for a second, off-line conversion step to the Windows Media format from an intermediate video format. This further reduces the workload burden on the operators.

Metadata. To reach the broadest audience within our community (e.g., HP Labs in Palo Alto, CA and Bristol, UK at present) with the media stored in the library, and to fully realize the effort involved in capturing the various lectures, seminars, and other presentations, metadata is collected for the individual media objects (i.e., streaming files). This feature enables subsequent indexing and searching of the media objects when uploaded to a media server. The kiosk requires that the user explicitly enter various metadata such as:

- Title
- Description
- Summary
- Author
- Copyright message

The following metadata is also collected automatically:

- Full name of user encoding the materials
- Time/Date
- Flag to indicate that the operator has confirmed adherence to copyright obligations
- Bit rate of media

The metadata is collected prior to encoding and stored in the Windows Media files themselves. The full name of the user who encoded the media is obtained seamlessly via their corporate NT account and an LDAP directory mapping NT accounts to Full Names (although that information is excluded from the printed example that follows). Please see Figure 1 below.

¹ <http://www.hp.com/workstations/ia32/xw8000/>

² <http://www.viewcast.com/products/osprey/osprey500.html>

³ <http://xinehq.de/>

⁴ <http://www.mplayerhq.hu/>

⁵ <http://windowsmedia.com/9series/DemoCenter/PlayBack.asp?page=6&lookup=PlayBack>

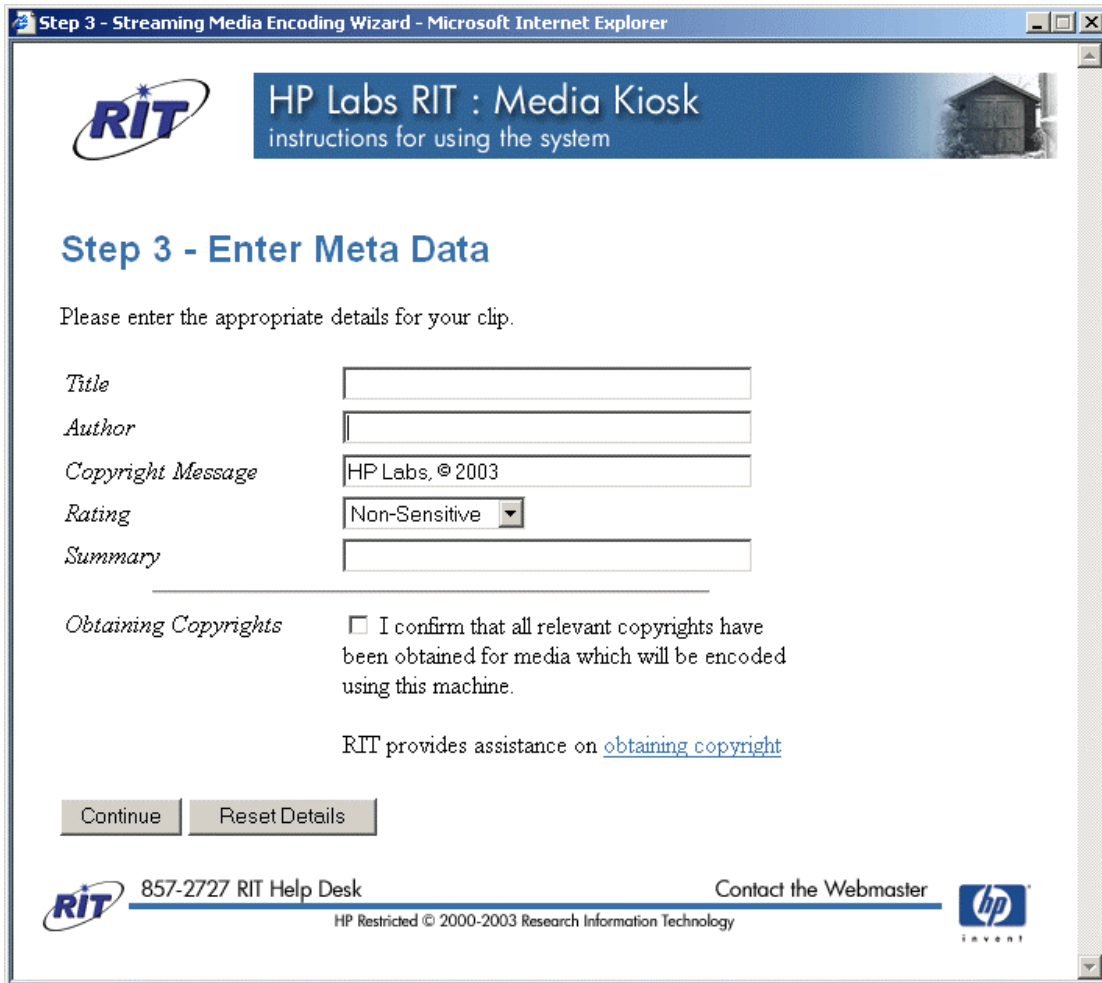


Figure 1

The Media Wizard that runs on the kiosk was described earlier. This wizard has a graphical user interface that includes instructions for the operator about the self-serve use. Moreover, it programmatically creates web pages that describe how to physically set up the VCR with the kiosk, what information is needed for the metadata, and legal requirements via provided forms. The GUI also provides feedback to the user as to progress. The wizard communicates with operators via Windows dialog boxes. There, they select content format, content placement, and error messages. As seen in Figure 2 below, users direct their files pre-specified locations. If they are not authorized to write to one of these places, other arrangements are made via an email (or call) to the Research Information Technology (RIT) help desk. Please see Figure 2 below.

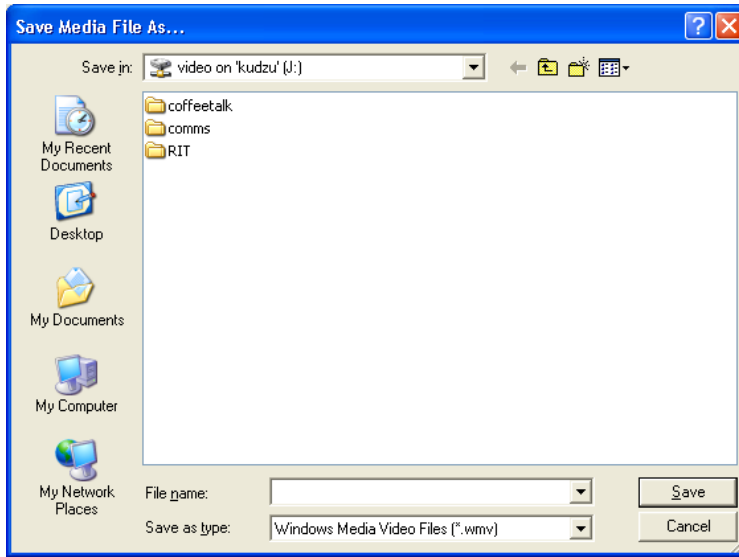


Figure 2

Users (self-serve kiosk operators) can select the target format of the media as seen in the dialog box below in Figure 3. While the typical selection is for intranet playback, at times content is necessary to be written primarily for CD-ROMs so viewers can take the materials along on a road trip where (1) connectivity or bandwidth may be limited or spotty; and (2) they still wish to review the talks or presentations. Please see Figure 3.

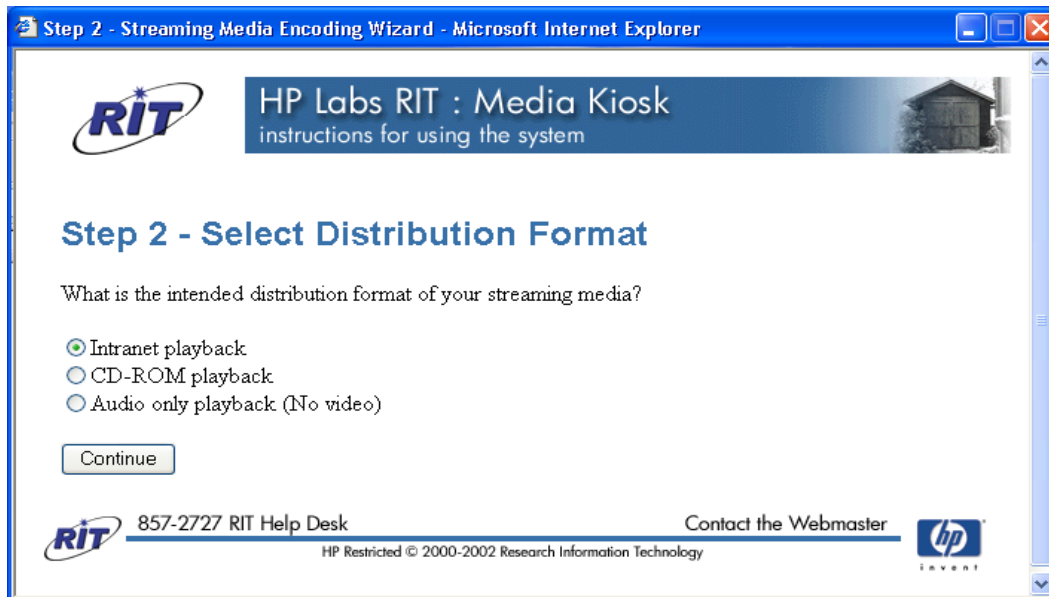


Figure 3

The service is also interruptible, should that be necessary. However, the software at present does not allow coding to continue from a previously interrupted place. In order to safeguard against accidental interruption, another dialog box is used as seen in Figure 4.

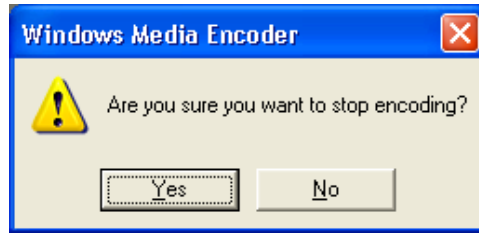


Figure 4

The encoding process on the capture kiosks is provided by the means of a VBScript which is approximately 2,250 lines long. This script utilizes objects provided by Internet Explorer⁶ (IE) to provide forms to the users and objects provided by Windows Media Encoder⁷ (WME) to provide the encoding interface. The WME interface provides the ability to programmatically control the options in the encoder such as the audio and video bit rates, the output destination, and many other features such as entry of the metadata, format density/destination, and control over the audio mixer. It is this that enables the user to encode the media, and make it available to the community, without needing to wade through a tangle of settings, options, and technical details.

The Graphical User Interface. A simple window is displayed to the users via the WME VB objects that enables the user to monitor the video output from the encoding session. The WME screen also provides the ability to monitor audio levels and adjust them via a slide bar to ensure optimum settings.

Uploading Streaming Files. The final step of the encoding process is to upload the media file to the media server where it is instantly incorporated into the media library. It is then readily available to end users. If a user wishes to upload the encoded media to the media server library they must first request an area of storage on the media server through a manual process. Server space is then allocated in the library hierarchy and then becomes available via a standard Windows share. This share is then mapped on the capture kiosk as part of the encoding process and users are able to choose where in the hierarchy they would like their media objects to appear, providing they have appropriate access privileges (e.g., employees in the Communications Department can write to their Comms shared space, but a casual user would not have such access privilege). The user is informed that the encoding process has terminated successfully both on screen and by an automatic email sent to their mail box. This message contains a link to their new streaming media object.

⁶ <http://www.microsoft.com/windows/ie/default.asp>

⁷ <http://www.microsoft.com/windows/windowsmedia/WM7/encoder.aspx>

The Streaming Servers

The servers are standard Dual 1GHz HP Netserver 2000R⁸ running Windows 2000 Server⁹ with approximately 200 Gb of storage capacity to host the media library. Each of the servers provides a Windows share in to the media library storage area, an NTFS folder which contains a hierarchy of folders relating to various departments in the organization, including one for a common space for any employee to encode then post streaming media objects.

The streaming servers run three key pieces of software with respect to the media library.

- Media Indexer
- IIS Web server
- Windows Media Services (WMS)

The media indexer is a custom piece of software that indexes all the Windows Media files stored in the media library, collecting the file name, the URL used to access it via the WMS server, and all the metadata stored within the file. This information is placed in a single media index on a SQL Server 2000 database hosted on another machine within the organization. This single index contains all the media from the various media servers hosted within HP Labs across various geographic locations.

A media library web application written in ASP is hosted on the IIS web server. This provides viewers with a way of discovering which library media are available. Each streaming server hosts an identical copy of the web application and provides the ability to access the content on any of the media servers within the organization from any of the servers' web applications.

Content can be accessed from two mechanisms: search and browse. The *search* facility provides a simple search which returns all media that matches a set of keywords and a more advanced search with which users can specify more detailed filters for each of the metadata fields. The *browse* facility allows the web application users to first select a server to browse and then move through the library hierarchy via a web representation of the files and folders on each server. Given the geographic nature of the organization, the separate servers intuitively make sense to viewers because they typically know where a talk was presented (e.g., in Bristol, UK) so that's the first server they'll browse, although the indexing and browsing functions do cover all servers. Once a file is identified, the resulting screen looks similar to the one in Figure 5 below. (Another aspect of the multiple server approach is to provide for effective replication, but that is outside the scope of this paper.)

⁸ <http://netserver.hp.com/products/info.asp?pid=lp2000r&sheet=specs&pid2=lp+2000r>

⁹ <http://www.microsoft.com/windows2000/server/evaluation/business/overview/default.asp>

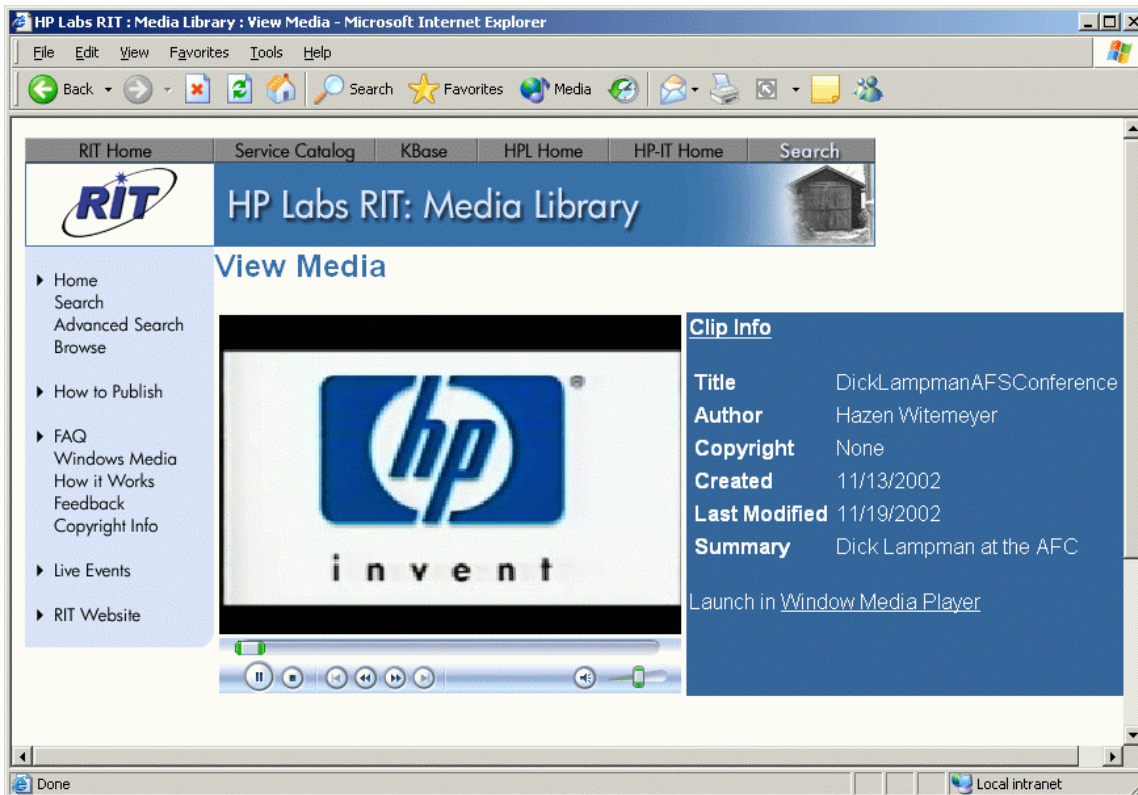


Figure 5

Each Windows Media file in the system has a dynamically generated associated web page allowing the user to view or listen to the file inside the browser and displaying the various metadata related to the file. The Windows Media ActiveX plug-in allows the file to be viewed inside Internet Explorer. Naturally, it is not displayed on non-Microsoft browsers which don't support the proprietary Active-X technologies. The media files can, however, be played through other client applications not limited to a single browser family. By providing a browser interface to the Windows Media files they can be easily indexed by intranet web crawlers. The summary metadata field is also included inside an HTML meta tag in all pages relating to a single piece of media. This improves the indexing of these pages as most web crawlers will recognize this tag as part of their indexing process. The media file specific page also contains an absolute link to the media file on the respective media server.

The servers run WMS configured as a single, on-demand uni-cast publishing point that is the media library storage area. WMS is shipped as standard with Windows 2000 Server edition. Once the server is configured with the publishing point there is no administration overhead to add content to the server. Media can be added to the library via the Windows share and accessed via the MMS protocol, as supported by WMS. If a file is not present it will return an error to the user's client.

The Display Clients

Viewers typically watch and listen to video streams on personal productivity computers (notebooks and desktops) that are running either XINE, mplayer for Linux or Windows Media Player. The only requirements of the display clients are that they are able to communicate using the MMS protocol and play streamed Windows Media (audio and video) .wmv format.¹⁰

On the Linux platform, evaluation of two different clients resulted in the selection of mplayer. The authors keep abreast of developments in the Open Source arena for other media players and upgrades. The other client, XINE, seemed to have a dependency on the quality of the stream, the local disk being the best, CD-ROM next, and the network being the worse, although both clients are in use. Using the MMS protocol caused XINE to occasionally drop the audio track, then it would hang. As a result of the evaluation, it was also recognized that mplayer has these advantages and attributes (The version examined was 0.9.13-fr1):

- The service delivers content over the network using the MMS protocol and mplayer supports that well.
- Mplayer played the content completely under the test conditions without hanging or problems with the audio or video.
- Performance over ISDN for mplayer is acceptable. Since there are a number of users who have such a connection this is important at present.
- A number of command-line options are available allowing the viewer to tailor to their preferences or circumstances.
- It appears that Linux distros are beginning to give a preferred place to mplayer.

This does not mean that XINE would never be used because an update or testing of another version of the software might fix the problem and acquire advantages over mplayer. With Open Source Software bug fixes and capabilities leapfrogging are not uncommon between various applications within a general area.

On the Windows platform the preferred client is Windows Media Player. Since Microsoft tunes their applications to work with their servers and that the WMP is embedded with the operating system, this is the logical choice from both a distribution and support point of view. The version must be Windows Media Player 7 or higher and the kiosk encoding decision is controlled by how the file is encoded. This enables virtually every Windows desktop at our site to be able to use our service. Currently, most of our desktops personal productivity platforms run Windows, although Linux desktop productivity platforms are on the rise.

¹⁰ <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwm/html/MMSFirewall.asp>, http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwm/html/optimize_web.asp, and <http://support.microsoft.com/default.aspx?scid=%2fsupport%2fmediaplayer%2fwmpptest%2fwmpptest.asp>

The Storage Challenges

Video generally produces large quantities of data which in turn creates numerous challenges which must be balanced when designing and implementing a system that manages this data. The main challenges are as follows:

- Storage costs
- Network bandwidth consumption
- Server load
- User experience

These factors needed to be balanced when choosing a bit rate for the video encoded by the capture kiosk. The choice of codec was fixed on the Windows Media 8 codecs, as supported by Windows Media Encoder 8 and Windows Media Player 7. Earlier versions of Windows Media Player are backward compatible through a mechanism that automatically downloads new codecs from the appropriate web site. This decision was based on two factors: (1) Windows Media Encoder 8 provides the advanced scripting facilities required by the capture kiosk; and, (2) the codecs were found to be very efficient at producing high quality video with low bit-rates, appropriate for the targeted delivery clients.

A lower bit rate produces smaller files for equivalent video sequences, this in turn reduces storage costs and server load, enabling a greater number of video streams to be supported simultaneously. Encoding video at a lower bit rate also reduces network bandwidth which keeps network support teams happy.

As bit rates increase so does the quality of the video that is encoded: the image is clearer and more fluid, giving a better end-user experience. Within HP Labs most desktop machines are directly connected to at least 100Mb networks, enabling high bandwidth streams to be delivered. However, many users may wish to view media from home or when on the road. From these locations the available bandwidth can be as low as 56k or less.

In order to deliver streams to all sites that are as useable as possible the ability of Windows Media 8 codecs to encode at multiple bit-rates in a single stream is utilized. For intranet playback streams are encoded at both 28Kbs and 160Kbps (audio and video combined), producing media files that are approximately 1Mb in size per minute of video. During playback the MMS Protocol, as implemented in the various display clients, provides facilities to report back to the server the currently available network bandwidth, which in turn enables the server to switch to a different stream if required. In practice it was found that network congestion issues became prohibitive to the use of higher bit-rates before storage limitations. The bit-rate settings were selected after an informal test and surveys at various sites from which the users may wish to access media.

Using The Service

The main goal of the service is to be an easy to use service: that is, anyone in our organization should be able to digitize content for widespread distribution, and anyone in the labs should be able to view the files whether they are using Windows or Linux on their desktops/laptops. By using the same authentication scheme as we use at our desktops, service is open to all. When writing to our servers for distribution, first time users are directed to contact the Research Information Technology media support team, who then creates a writable folder on the appropriate server. From then on, users may come up to the kiosk, digitize the content and expect the final product to be available, searchable by the search tool and viewable from the web pages on the Windows desktop or via the aforementioned client media players for either Linux or Windows desktops.

Simplicity is a key goal. There are really two parts to this service where we have an authoring part and then a distribution part. Since the kiosk as mentioned before is automated, there is no need to copy files, open up network shares, and other manual processes. Also linking the metadata of the content to the streaming files themselves is not a burden on the operator. Links are directly presented on a web page when executing a file browse. Or, the link may be copied by the author and distributed via e-mail to prospective viewers as an announcement. In this way, e-mail communication is the only other step needed to alert viewers to the content. Seconds after the digitizing, distribution throughout the community is available.

Training on how to use the service is communicated in a variety of ways. These include the web pages, the Media Wizard's auto-instruction walk-through, and personal one-on-one instruction of those who request it. The web pages and the Media Wizard are sufficient in most cases due to ease of use. Anecdotal evidence of this is found in that no one has ever called back to say "show me that again." Once a person has been taught, e-mail messages typically follow stating that the content for a presentation or a talk is now online.

Communicating The Service

As a large distributed organization, and the ever present email flood, it is challenging to communicate the availability of both the digitizing service as well as the programs that are available. Multiple channels are used, and these include the ongoing involvement of the Communications Department whose email announcements of important programs, word of mouth, and the growing familiarity of browsing the servers for content. Additionally, an article in the Labs' Research IT Group's newsletter introduced the service and programs and pointed to a set of web pages. Also, instructional one-on-ones continue to be employed.

While the main purpose was to deliver a self-serve service to the Communications Department, the project focused equally on meeting the needs of the entire Labs community. However, not all geographic locations yet have kiosks on site enabling their ease of digitizing tapes, but anyone within the Labs community can access the programs. When

programs from the capture kiosks were ready, e-mail notifications were sent out with links to the streaming media content. Viewers started to view the talks online.

The wizard also has a Flash movie sequence that instructs users on how to start the encoding process of tape in the VCR. Thus the Media Wizard plays a crucial part in communicating the service for the user who is the author of the content files. The design called for the Media Wizard to be part of the service itself.

HP Labs has a Research Information Technology (RIT) newsletter that conveys announcements, articles, and other useful information to our research community. The capture kiosk developers published an article introducing the service with instructions on how to use it. This was for the authoring side of the service. It also alerted users that they would be able to view talks online had they been unable to attend them previously. Included were relevant web links that explained the service in greater detail.

RIT also provides instructional one-on-one training sessions for authors. Although the kiosk instructions are sufficient, the personal touch goes a long way towards making users comfortable and being able to answer their spontaneous questions. In every case, each author who has received personal training commented on how easy the process was. Some of the trained authors have now trained others.

Future Directions

One of the lessons learned is that if you build it, they *might* come. In this case they came, and more came. This underscores the value of prototyping. The Streaming Media Capture and Delivery service eased itself into our infrastructure without visible fanfare. Authors who are trained, just went up to the kiosk and created content. The Communications Department just sends the URL and people watch the presentations. It is pervasive now and, having become commonplace, people expect it. This is interesting because this was a prototype service. Another lesson learned is that the value of making staff personnel available for face-to-face training, more for the social context than an instructional one, but this nevertheless fostered the acceptance of the prototype.

Work in this area should be continued. The issue of editing the content has not been addressed, but there are many solutions available. Continued effort not only as the Windows Media platform evolves but also for other platforms as well, notable Linux. This model could be used, for instance to setup a service using the Helix DNA components that run on Linux as well as others that claim to deliver universal formats, including Windows Media. Also work on modifying the current Media Wizard to do one button live encoding, multi-/uni-casting, archiving and indexing is on the horizon.

Similarly, live capture is of increasing importance, however the cost of the video boards is significant. We're exploring how to use more powerful machines (e.g., twin processors) and FireWire connectivity to bypass a dedicated board. As the level of computing power continues to increase, this has the potential to simplify the overall capture kiosks and turn

it instead to a software solution downloadable by anyone instead of a standalone hardware/software kit. This will enable more of our Labs in disparate geographic locations to capture their own talks and subsequently make them available for the worldwide HP Labs' community as well as interested HP colleagues around the globe.

Also of note are architectures that could more effectively capture live video on a site-wide basis, that is, to route video signals from the various meeting rooms to a farm of encoders. This notion does call to mind various needed modifications to the general physical plant of the buildings in terms of the routing of video and/or networking cables, or the use of portable, prototyped appliances configured explicitly for the capture and encoding of video streams. With this, we immediately start to see application for utility type computing.

Also under consideration are possible integration with various collaborative tools to better support encoding of content for "meetings-on-demand" or to feed-forward information from a meeting in one time zone to a subsequent meeting in another. It is possible that streaming media capture eventually will be a client capability on all desktops, although this is not a current plan. Also involved are various management and organizational aspects as to whether a person needs to be designated as the "camera operator" in order to capture the collaborative session. A new meaning to phrase: "please take notes" might emerge from this approach. The capacity to index and cross-correlate video streams with documents that are shown during a presentation via timestamps or other index markers is a continuing area of research. We can also see that high usage frequency at the small meeting level leads to considering such architectures as the Utility Data Center to support multiple feeds and distinct encoding streams, thus enabling diverse users to revisit their conversations and whiteboard squiggles from the captured media objects.

A topic we've not yet addressed is the realm of content lifecycle. Some talks certainly endure - ones that discuss a controversial book or theory, a presentation by a Nobel Laureate, or the like will be viewed for years to come. Yet other talks may be only concerned with current commercial research strategies, and while critically important in the short term, may well be of no use in as little as 90 days. These two examples raise the question of how to identify and then manage the lifecycle of the various content objects that can be captured by the encoding kiosks. No decisions on these topics have yet been made, other than to observe the frequency that various files are viewed.

As the service becomes more formalized and adopted as a "production" service, it is necessary in a large organization to engage the user support Help Desks with guidelines as to how to handle support calls (e.g., a kiosk system crashes or someone kicked out the wall plug), and prepare various help files and usage scenarios. Similarly, live demonstrations to groups of potential kiosk users are useful to foster adoption.

Conclusion

Our experience at HP Labs has shown that with a relatively small investment in programming time and hardware purchases, valuable talks and presentations can be made available to our

user population unbounded by either time or space. People can now watch the streams at their convenience irrespective of the original scheduled time of the talk or presentation. Moreover, streams can be viewed at various locations (e.g., when people are working at home) provided that sufficient bandwidth is available. This useful service has become commonplace and the authors expect that as the aforementioned future directions are pursued, the convenience will continue to improve and usage become more widespread. Multimedia technology has thus become a positive, contributing factor in the intellectual environment of HP Labs.